

## **REMARKS**

Claims 1-12, 14-37 and 39-42 were presented for examination and were rejected. Applicants thank the Examiner for examination of the claims pending in this application and addresses Examiner's comments below.

Reconsideration of the application in view of the above amendments and the following remarks is respectfully requested.

### **Claim Rejections Under 35 U.S.C. §§ 102 and 103**

In paragraph 2 on page 2 of the Office Action, the Examiner rejected claims 1-3, 6-12, 14-18, 35-37 and 39 under 35 U.S.C. § 102(e) as being anticipated by Rubin et al ("Rubin", US 200210099552). In particular, the Examiner contends that Rubin discloses an apparatus for direct annotation of objects including a display device; an audio input device; and a direct annotation creation module coupled to receive an input audio signal from the audio input device and to receive a reference to a location within an image from the display device (Figure 4, [0048] lines 1-6, [0066] lines 1-3), the direct annotation creation module, in response to receiving the input audio signal and the reference to the location within the image (Figure 4), automatically creating an annotation object, independent from the image, that associates the input audio signal with the location.

Furthermore, in paragraph 3 on page 8 of the Office Action, the Examiner rejected claims 4, 5, 19-34, 40 and 42 under 35 U.S.C. § 103(a) as being unpatentable over Rubin et al in view of Mitchell et al ("Mitchell", US 5,857,099). In particular, the Examiner contends that Rubin fails to distinctly point out details of the audio voice recognition technology.

However, Mitchell discloses the apparatus of claim 1 further comprising: an audio vocabulary storage for storing a plurality of audio signals and corresponding text strings (Mitchell, Figure 2 item 20); an audio vocabulary comparison module, wherein the direct annotation creation module uses text strings found by the audio vocabulary comparison module to create the audio annotation (Column 5 lines 25-65). Therefore it would have been obvious to an artisan at the time of the invention to combine the vocabulary-based audio conversion of Mitchell with the system of Rubin. Motivation to do so would have been to provide a convenient way to convert voice to text (Rubin, [0097] lines 1-9).

Applicants have amended claims 1, 4, 5, 7-10, 14, 15, 26, 35 and 40 to more clearly define the claimed invention.

Claim 1 has been amended for consistency of the terms used and now recites:

a storage device for storing a ***plurality of different visual notations***; and  
a direct annotation creation module coupled to receive the audio signal from the audio input device and to receive a reference to a location within an image on the display device, the direct annotation creation module, in response to receiving the audio signal and the reference to the location within the image, automatically creating an annotation object, independent from the image, that associates the input audio signal, the location ***and one of the plurality of different visual notations***.  
[Emphasis Added.]

Independent claims 7-10, 26, 35 and 40 have been similarly amended to include language to clarify that the present invention creates a visual notation or label that is different from other notations related to annotations to the image to indicate: the presence of an annotation and subject matter of the annotation.

Applicants respectfully submit that the independent claims 1, 7-10, 26, 35 and 40, as amendment are patentably distinct over Rubin et al, alone or in combination with Mitchell.

The claimed invention has been amended to specify that each annotation includes *one of the plurality of different visual notations*, and that that one different visual notation is displayed in some instances. This claimed feature of the present invention is particularly advantageous because the claimed invention allows the user to automatically add an annotation AND to identify or set apart each annotation from other annotations. Moreover, the same visual notation can be used for the same subject matter to create groupings or sets, such as when there is an annotation about the same person in different images. The ability to provide different visual notations or labels is very different from anything disclosed by Rubin et al, or Mitchell. This differs dramatically from the cited art where the label is simply the same generic audio icon for all annotations. For example, in the claimed invention the label might be "Steve" for one location and "Dan" for another as in Figure 11C. These same labels would be used by the system on multiple additional images where those individuals were in the images.

The difference between the claimed invention and the prior art becomes even clearer and is set forth in greater detail in the dependent claims. For example, certain dependent claims such as claims also recited “vocabulary storage” or a “vocabulary comparison.” The claimed invention is directed to the comparison of audio input to stored exemplars that have corresponding “visual notations” or labels. The claimed invention compares the input with stored exemplars, labels the location of the input with the stored exemplar, and then links the input to the visual notation. Furthermore, if there is no sufficiently good match for the input, the claimed invention gives the user the option of creating a new visual notation and associating the visual notation with an exemplar. In one embodiment, that is not full-on speech recognition, but comparison between an input and stored exemplars. Several

techniques for doing this exist in the audio domain such as a kind of nearest neighbor of matching. Assuming a matching exemplar is found, the claimed invention labels the input with the visual notation associated with that exemplar. The claimed invention also maintains links between the visual notation and all such annotations that is useful for searching/navigating as well as improving the recognition accuracy. Finally, other claims specify that if a good match for an annotation is not found, the user is given the option to either create a new exemplar (and label) or pick an existing exemplar. In the second case, the audio input is used as an additional training point for that exemplar (meaning the accuracy in recognizing the input will increase for the next time that similar audio input is encountered.)

Applicants submit that the claimed invention is not taught or suggested by the art of record. Rubin has no teaching or suggestion to use visual notations of different types. Rubin consistently use the same icon to show an audio annotation. Thus, the user in the prior art cannot determine anything about the annotation other than it exists from the display of the annotation. Furthermore, Mitchell does not provide any teaching or suggestion to remove this deficiency. Mitchell is absent any teaching about labeling or different visual notations for annotations. Moreover, Mitchell does teach or suggest comparison of visual notations and exemplars. Thus, the present invention is advantageous over the prior art in at least two significant respects. First, the use of a visual notation from a plurality of different visual notations allows the user to distinguish and group annotations such as being related to a particular person or topic of interest. Second, the existence of such different visual notation from a plurality of different visual notations, allows all the annotations related to a visual notation to be search and grouped. Finally, the use such different visual notation from a

plurality of different visual notations can be applied to improve the search results and offer suggestions for labeling to the user. None of these advantages are provided by the prior arts use of a single icon for all annotations. Therefore, Applicants submit that the claims as now amended are patentably distinct over Rubin and Mitchell, alone or in combination.

Based on the above Amendment and Remarks, Applicants respectfully submit that for at least these reasons independent claims 1, 7-10, 26, 35 and 40, are patentably distinguishable over Rubin and Mitchell, both alone and in combination. Therefore, Applicants respectfully request that Examiner reconsider the rejections, and withdraw them.

Claims 2-6 are dependent on claim 1, claims 11-12 and 14-25 are dependent on claim 10, claims 27-34 are dependent on claim 26, claims 36-39 are dependent on claim 35, and claims 41-42 are dependent on claim 40. Each of these dependent claims recites other patentable features some of which have been discussed above. Thus, all arguments advanced above with respect to independent claims 1, 7-10, 26, 35 and 40, are hereby incorporated so as to apply to claims 2-6, 11-12 and 14-25, 27-34, 36-39 and 41-42, respectively

### **CONCLUSION**

In sum, Applicants respectfully submit that claims 1-12, 14-37, and 39-42, as presented herein, are patentably distinguishable over all of the art of record. Therefore, Applicants request reconsideration of the basis for the rejections to these claims and request allowance of the claims.

In addition, Applicants respectfully invite Examiner to contact Applicants' representative at the number provided below if Examiner believes it will help expedite

furtherance of this application.

Respectfully Submitted,  
GREGORY J. WOLFF, et al.

Date: July 24, 2006

By: /Greg T. Sueoka/  
Greg T. Sueoka, Reg. No.: 33,800  
Fenwick & West LLP  
Silicon Valley Center  
801 California Street  
Mountain View, CA 94041  
Tel.: (650) 335-7194  
Fax: (650) 938-5200

20412/06364/DOCS/1604069.3